# Evaluating the impact of a two-stage multivariate data cleansing approach to improve to the performance of machine learning classifiers: a case study in human activity recognition

Autores

Dionicio Neira-Rodado, Chris Nugent, Ian Cleland, Javier Velasquez, Amelec Viloria

Abstract

Human activity recognition (HAR) is a popular field of study. The outcomes of the projects in this area have the potential to impact on the quality of life of people with conditions such as dementia. HAR is focused primarily on applying machine learning classifiers on data from low level sensors such as accelerometers. The performance of these classifiers can be improved through an adequate training process. In order to improve the training process, multivariate outlier detection was used in order to improve the quality of data in the training set and, subsequently, performance of the classifier. The impact of the technique was evaluated with KNN and random forest (RF) classifiers. In the case of KNN, the performance of the classifier was improved from 55.9% to 63.59%.

Palabras clave

HAR, dataset quality, machine learning, multivariate analysis.